D1979 | EGU2020-2473 SSS5.2

Mechanisms of soil organic matter stabilization and carbon sequestration

Spatially heterogeneous mechanisms underlying soil carbon sequestration as revealed via big data-driven Earth system modelling and deep learning

Yiqi Luo, Northern Arizona University, United States

Feng Tao, Xiaomeng Huang, Tsinghua University, China

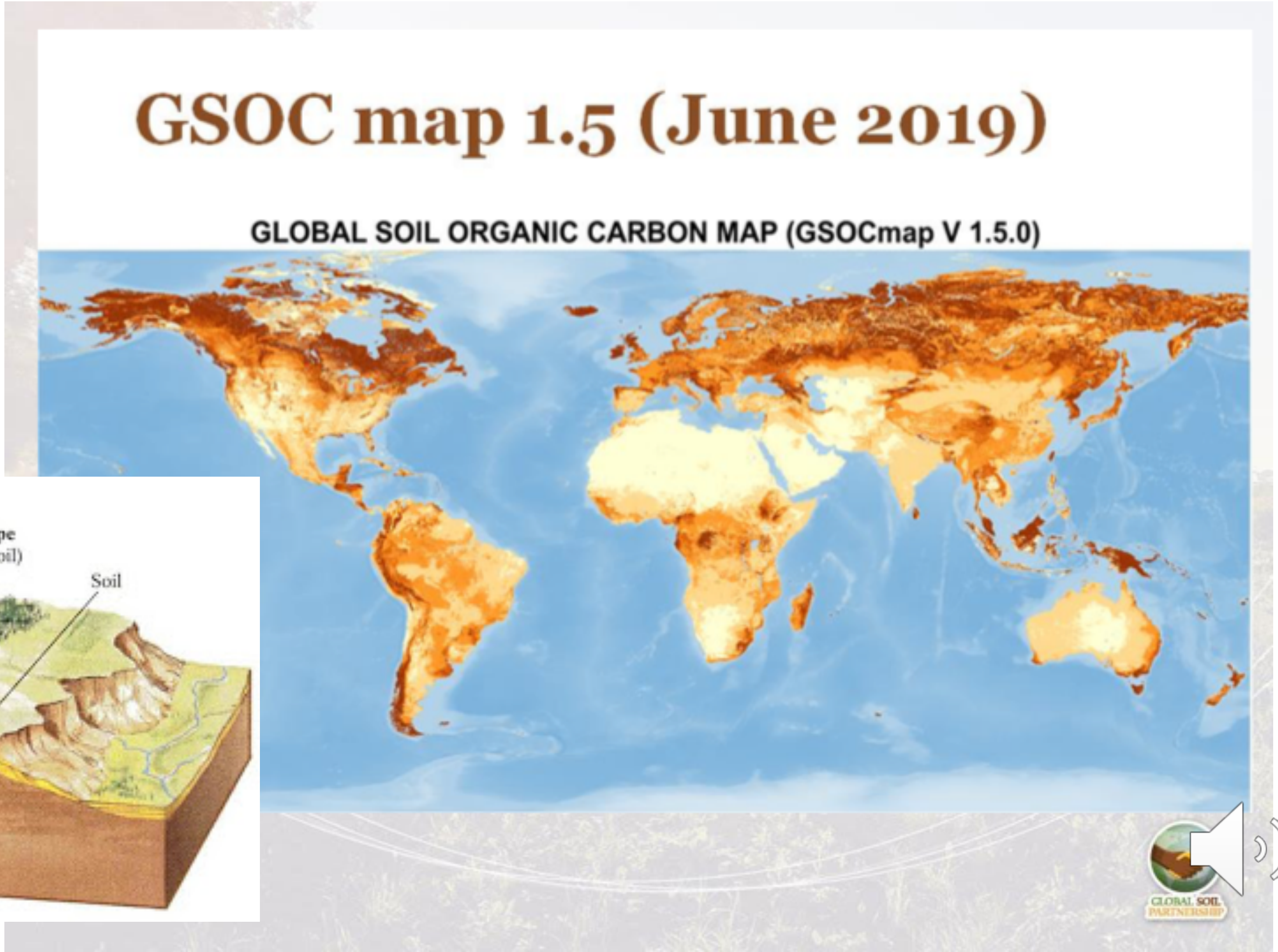
Yiqi.Luo@nau.edu

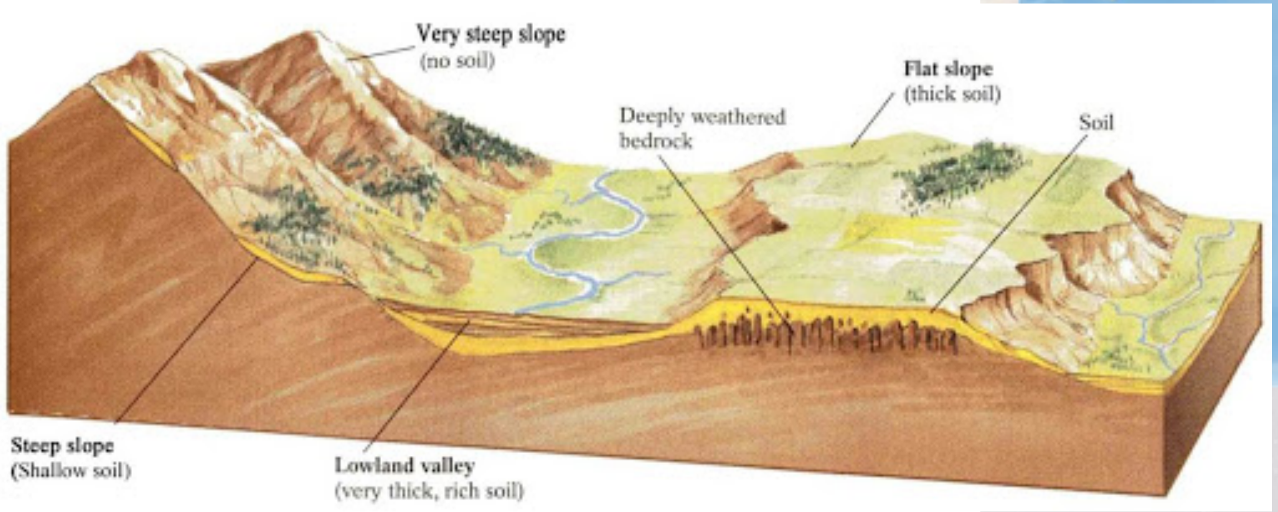
http://www2.nau.edu/luo-lab/?home

Center for Ecosystem Science and Society at Northern Arizona University

EGU virtual meeting, Wed, 06 May, 14:00–15:45, 2020

Soil heterogeneity at all scales from local pits to the globe

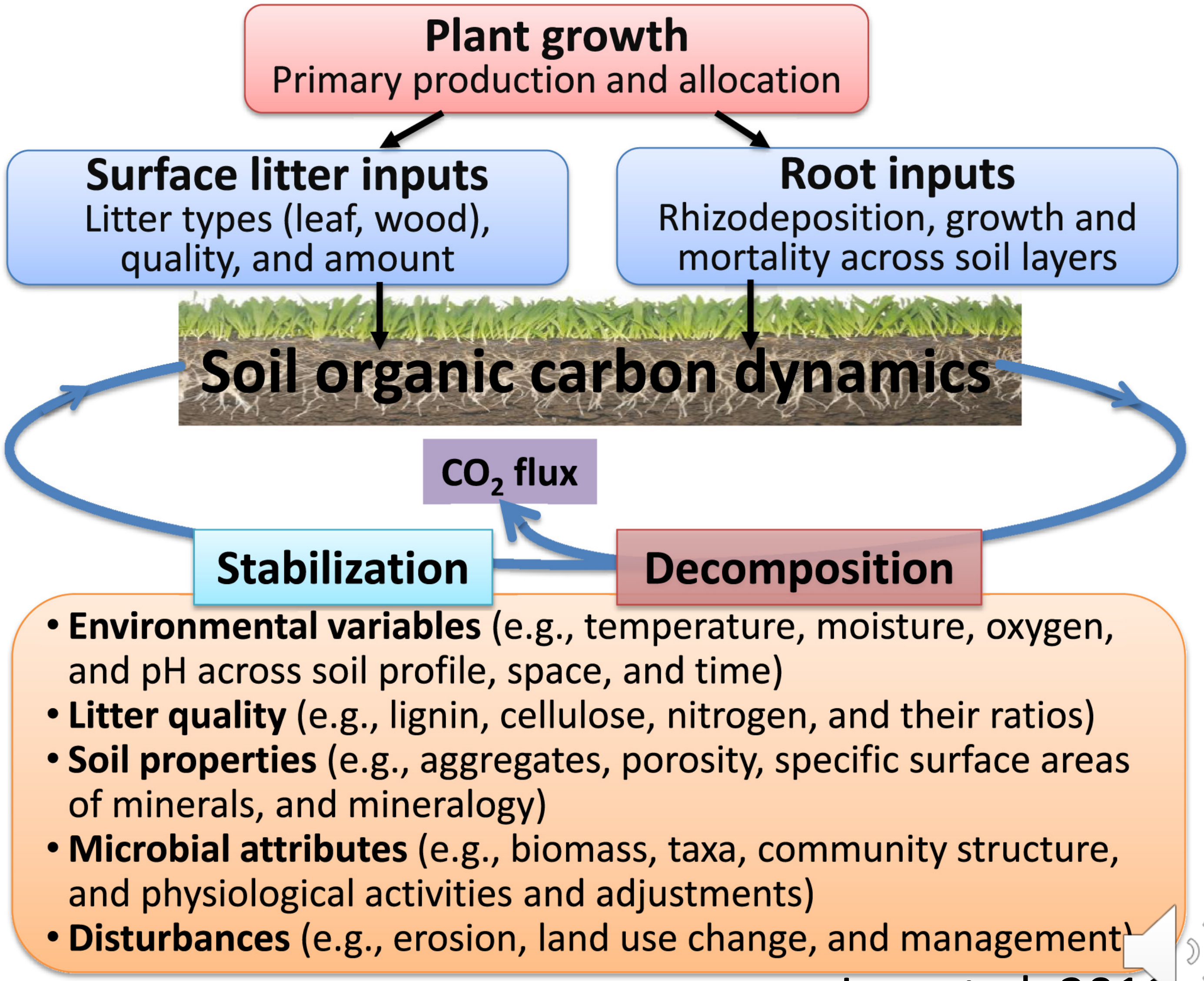




What determine the soil heterogeneity?

Empirical studies

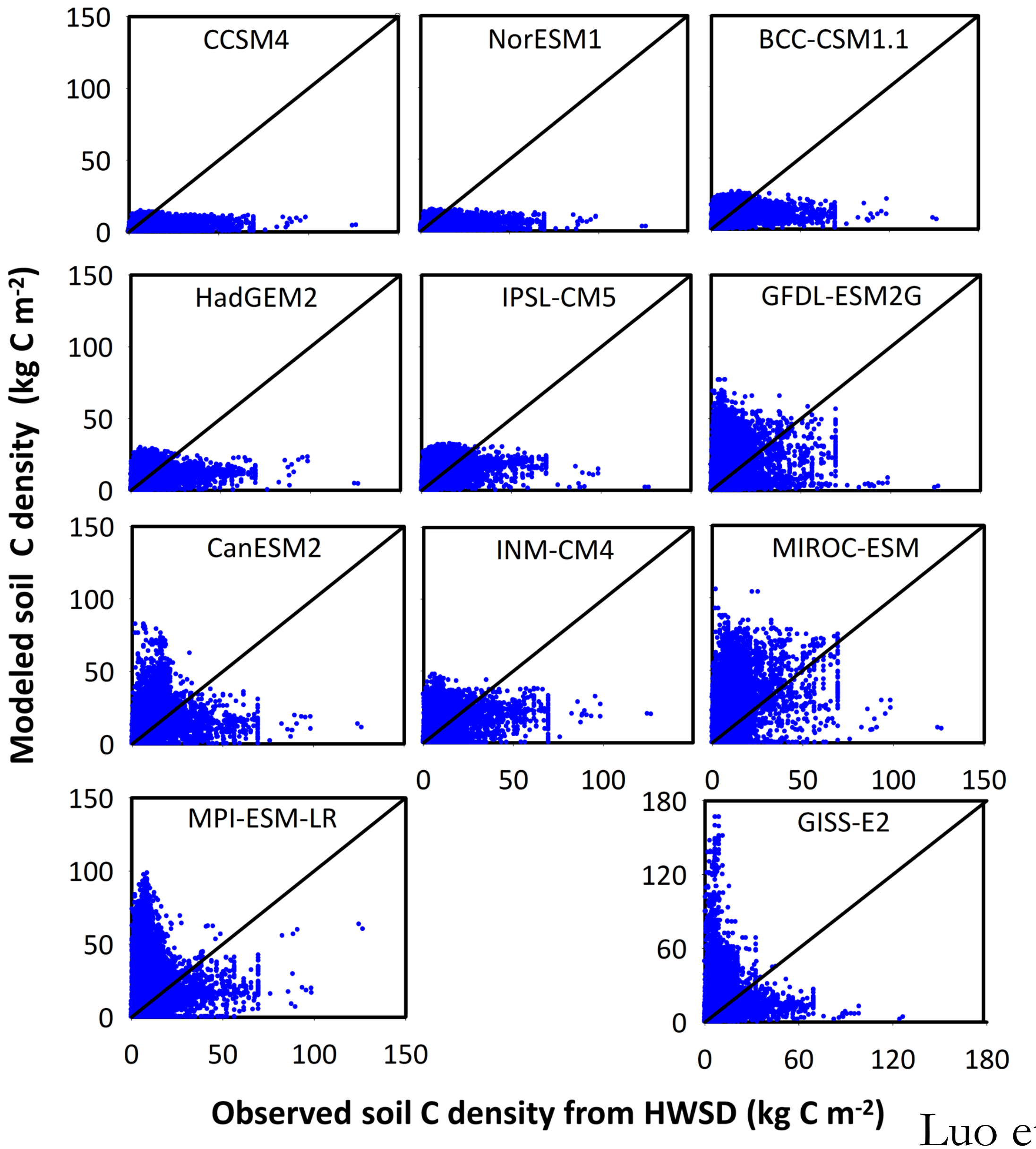
- ✓ suggest that many processes and factors can cause soil carbon heterogeneity
- ✓ have not identified key mechanisms over large scales



Luo et al. 2016

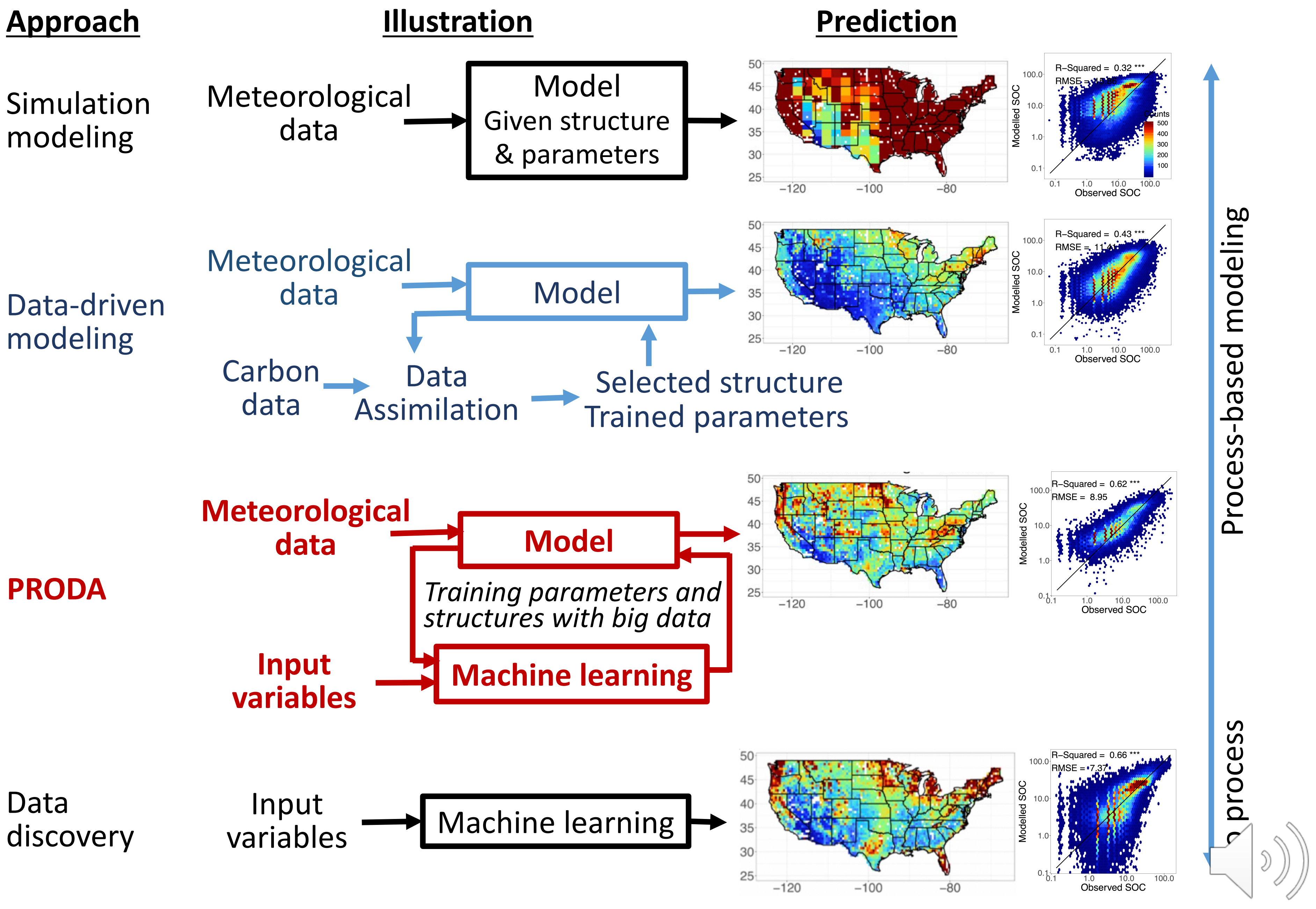
Model output for CMIP5

Models mostly
- ✓ use environmental scales (e.g., temperature and moisture) to account for soil carbon heterogeneity
- ✓ can not predict spatial heterogeneity well

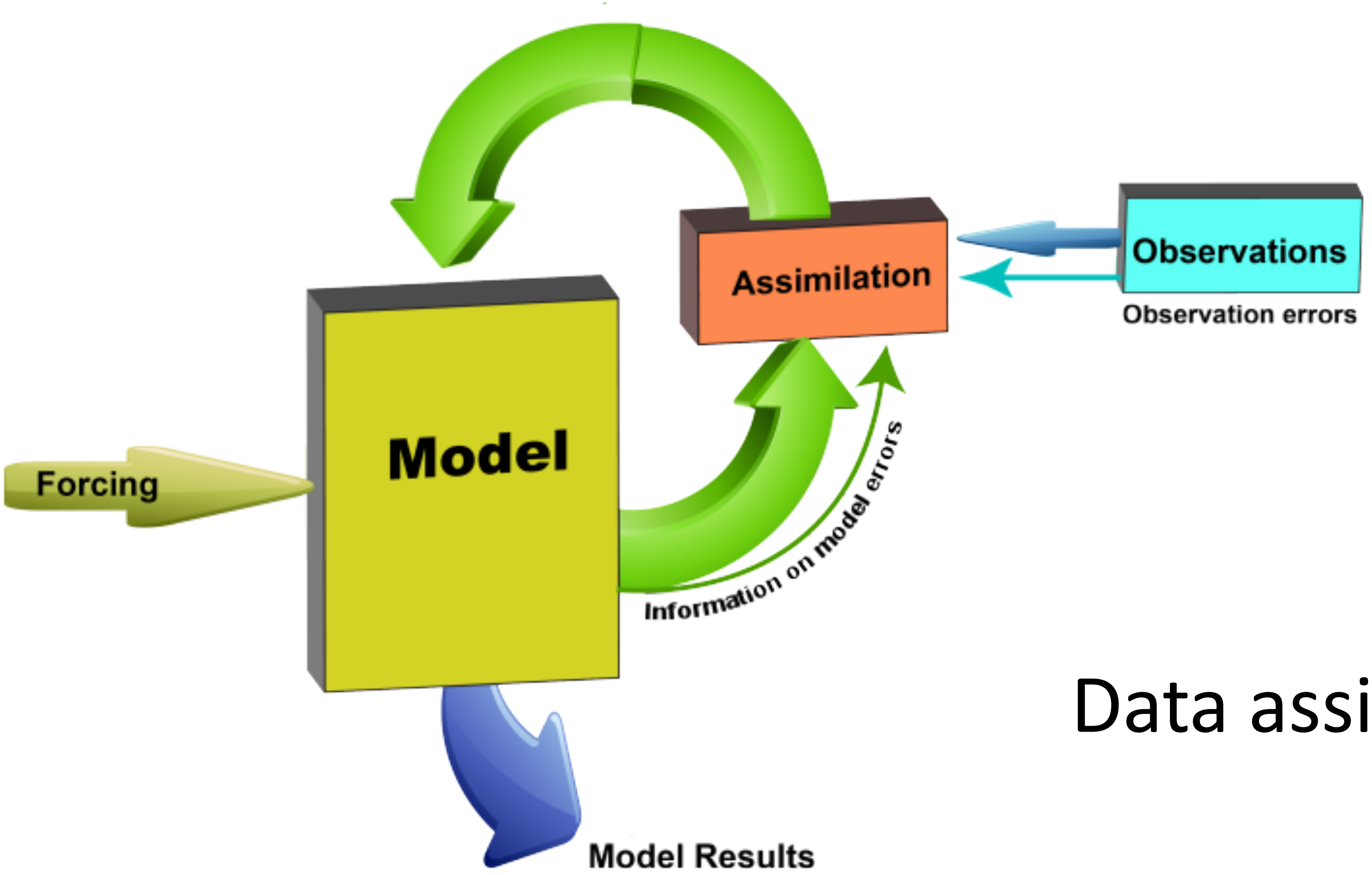


Luo et al. 2015

Methods


Approach
Illustration
Prediction
Simulation modeling
Meteorological data
Model
Given structure & parameters
Data-driven modeling
Meteorological data
Model
Carbon data
Data Assimilation
Selected structure
Trained parameters
PRODA
Meteorological data
Model
Training parameters and structures with big data
Input variables
Machine learning
Data discovery
Input variables
Machine learning
R–Squared = 0.32 ***
R–Squared = 0.43 ***
R–Squared = 0.62 ***
RMSE = 8.95
R–Squared = 0.66 ***
RMSE = 7.37
Modelled SOC
Observed SOC
Process-based modeling

Data-driven modeling



Data assimilation

Model

Matrix equation of CLM4.5

$$\boxed{\frac{d\boldsymbol{X}(\boldsymbol{t})}{d\boldsymbol{t}} = \boldsymbol{B}(\boldsymbol{t})\boldsymbol{I}(\boldsymbol{t}) - \boldsymbol{A}\boldsymbol{\xi}(\boldsymbol{t})\boldsymbol{K}\boldsymbol{X}(\boldsymbol{t}) - \boldsymbol{V}(\boldsymbol{t})\boldsymbol{X}(\boldsymbol{t})}$$

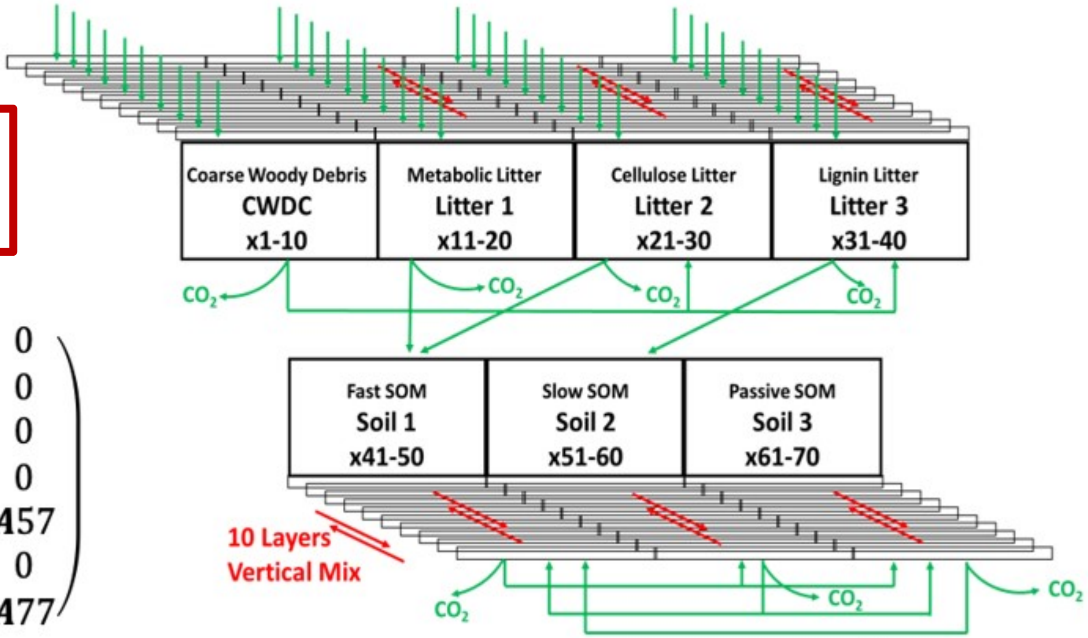
$$X(t) = (X_1(t), X_2(t), X_3(t), \dots, X_{70}(t))^T$$

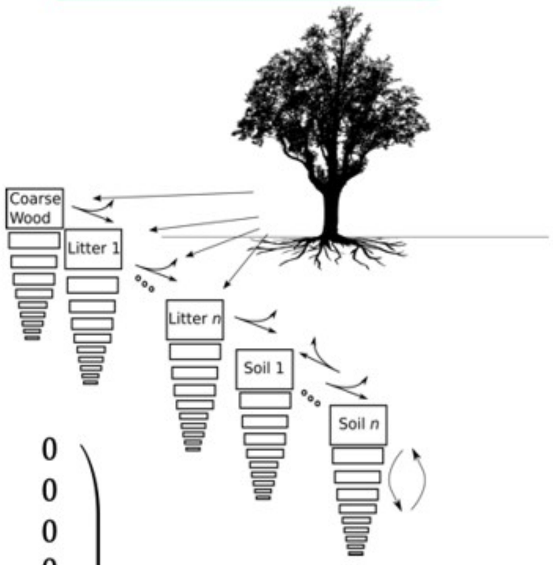
$$\mathbf{A} = \begin{pmatrix} A11 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & A22 & 0 & 0 & 0 & 0 & 0 \\ A31 & 0 & A33 & 0 & 0 & 0 & 0 \\ A41 & 0 & 0 & A44 & 0 & 0 & 0 \\ 0 & A52 & A53 & 0 & A55 & A56 & A57 \\ 0 & 0 & 0 & A64 & A65 & A66 & 0 \\ 0 & 0 & 0 & 0 & A75 & A76 & A77 \end{pmatrix}$$

$$\mathbf{A}_{31} = diag(-f_{31}, -f_{31}, -f_{31}, -f_{31}, -f_{31}, -f_{31}, -f_{31}, -f_{31}, -f_{31}, -f_{31})$$

$$\mathbf{V(t)} = \begin{pmatrix} V11 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & V22(t) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & V33(t) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & V44(t) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & V55(t) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & V66(t) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & V77(t) \end{pmatrix}$$

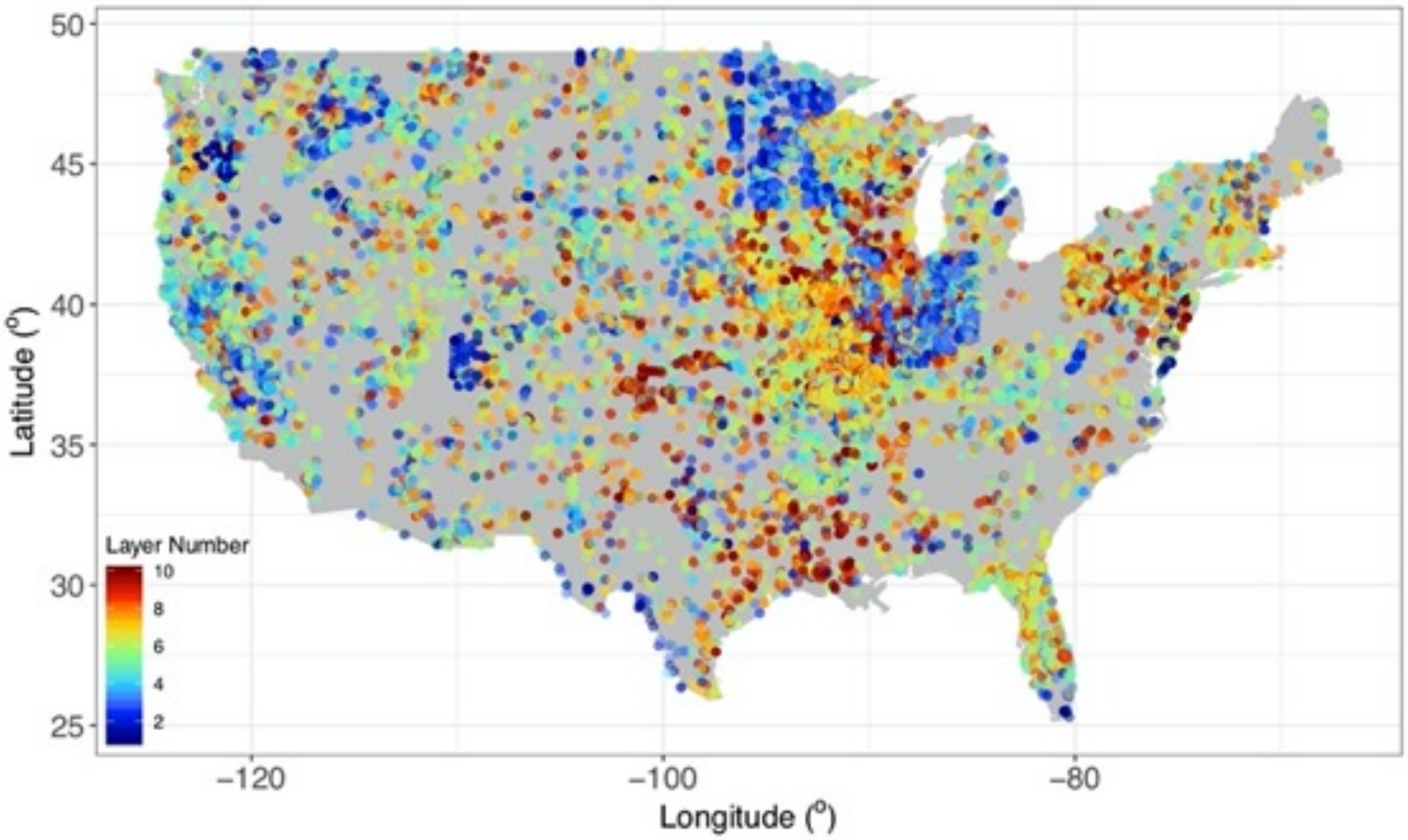
$$V22 = diag(z_1, z_2, \dots, z_{10})^{-1} \begin{pmatrix} g_1 & -g_1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -h_2 & h_2 + g_2 & -g_2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -h_3 & h_3 + g_3 & -g_3 & \cdots & 0 & 0 & 0 \\ 0 & 0 & -h_4 & h_4 + g_4 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & h_8 + g_8 & -g_8 & 0 \\ 0 & 0 & 0 & 0 & \cdots & -h_9 & h_9 + g_9 & -g_9 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -h_{10} & h_{10} \end{pmatrix}$$





Huang et al. 2018
Global Change Biology

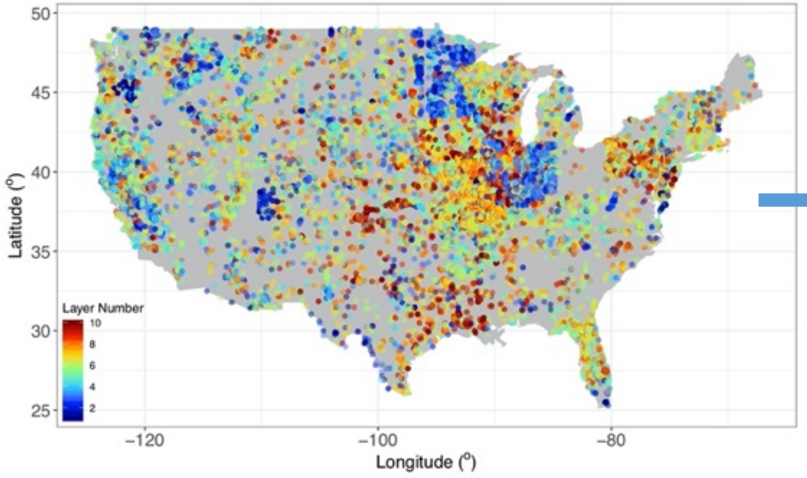
Data: >24,000 vertical profiles in US continent



Data-driven modeling

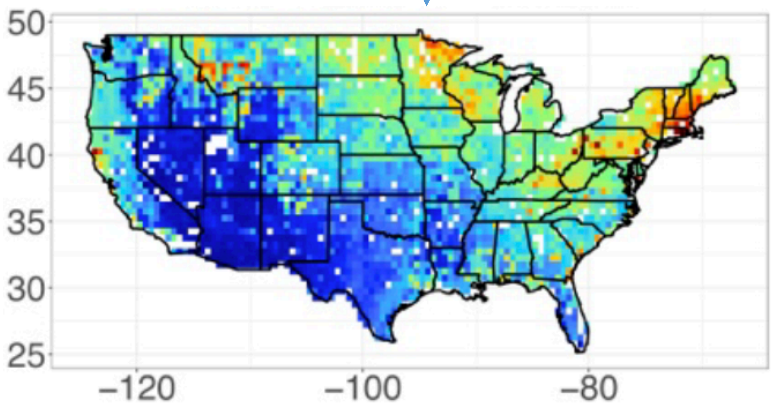
Meteorological data

$$\frac{d\boldsymbol{X}(t)}{dt} = \boldsymbol{B}u(t) - \boldsymbol{A}\xi(t)\boldsymbol{K}\boldsymbol{X}(t) - \boldsymbol{V}(t)\boldsymbol{X}(t)$$



Data Assimilation

Trained parameters



PRODA:
PROcess-guided machine learning and DAta-driven modeling

PRODA

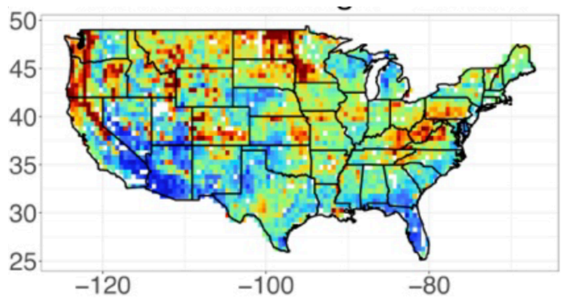
Meteorological data

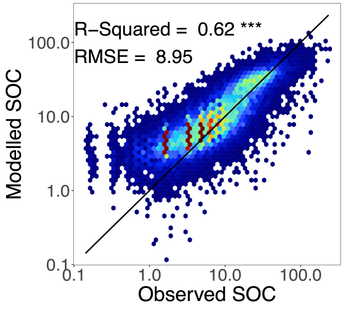
Input variables

Model

Training parameters and structures with big data

Machine learning





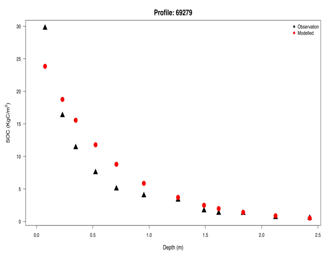
Meteorological data

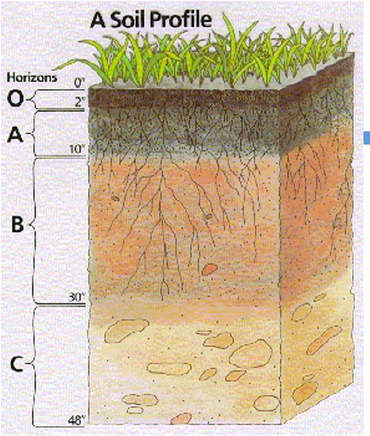
$$\frac{d\boldsymbol{X}(t)}{dt} = \boldsymbol{B}u(t) - \boldsymbol{A}\xi(t)\boldsymbol{K}\boldsymbol{X}(t) - \boldsymbol{V}(t)\boldsymbol{X}(t)$$

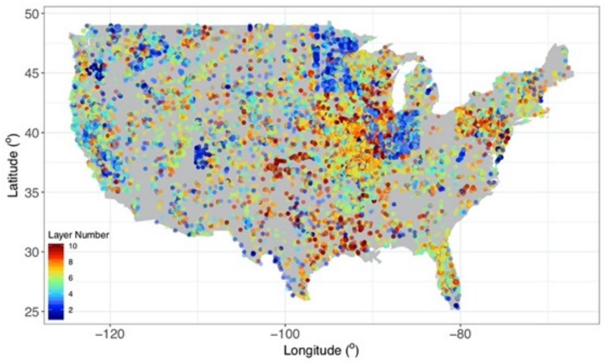
Prediction

Predicted parameters

~24,000 vertical profiles in US continent



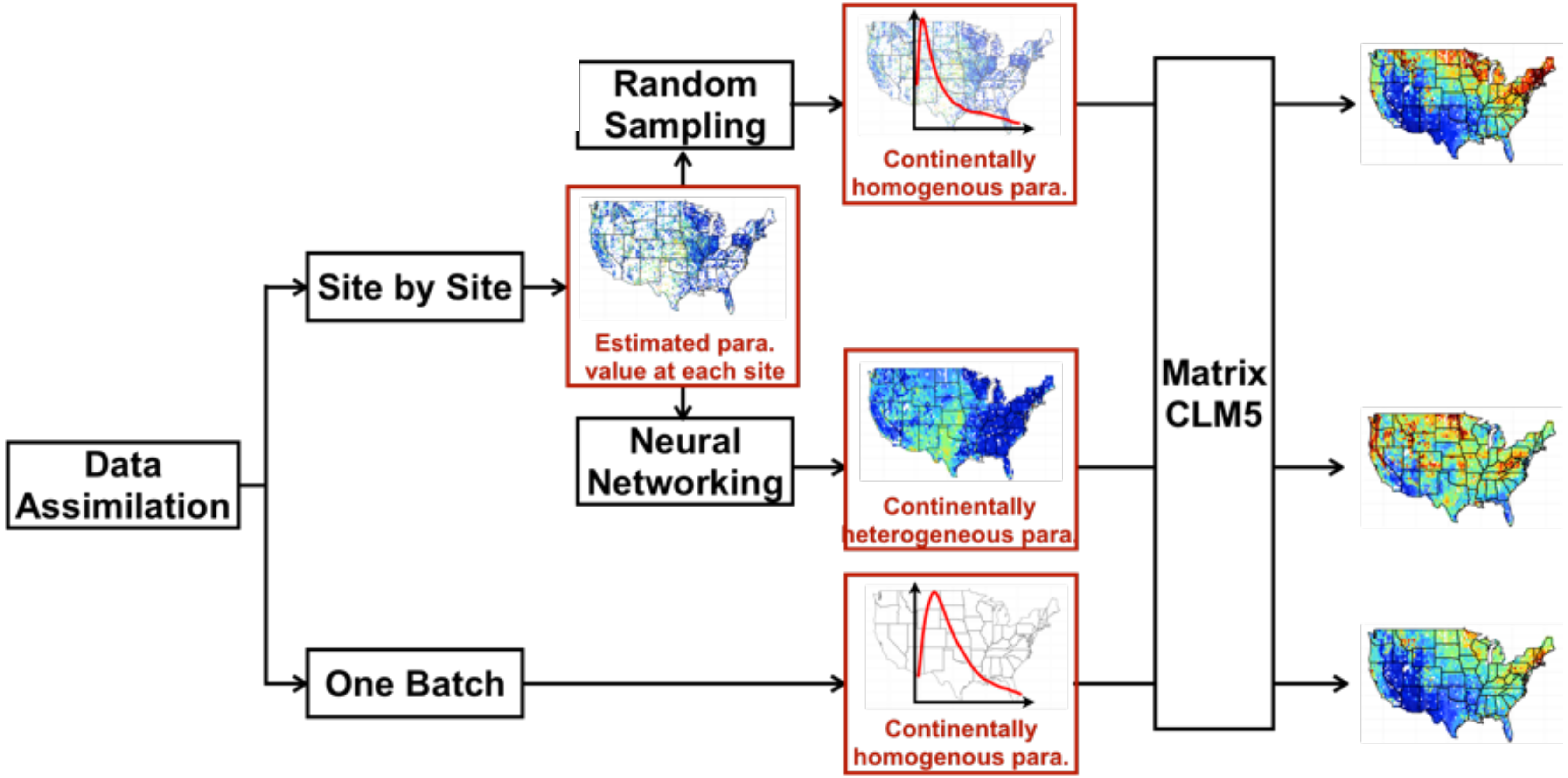


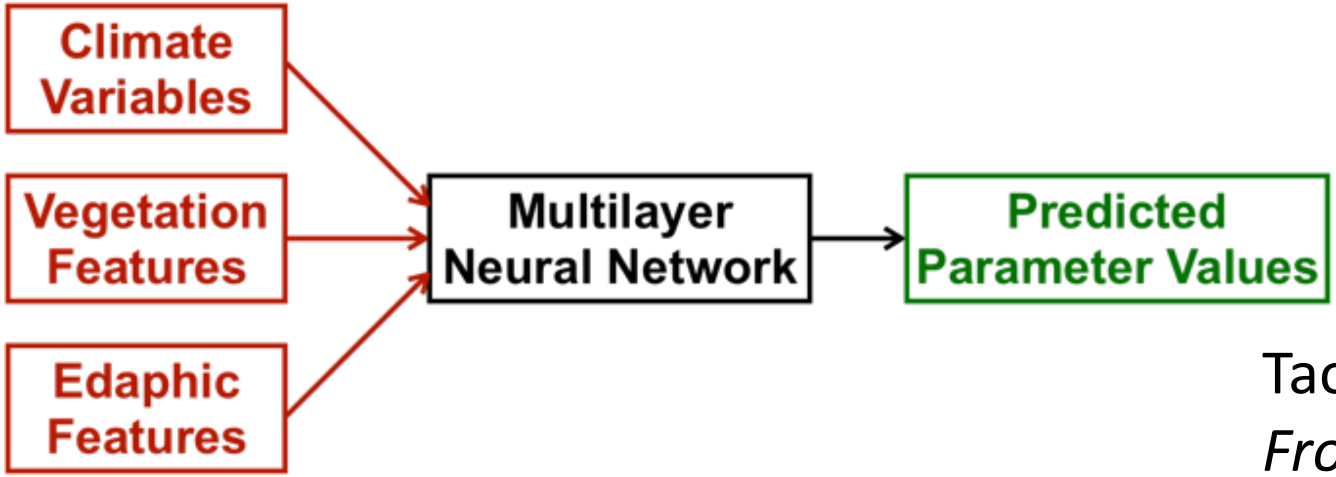


Site-by-site Data Assimilation

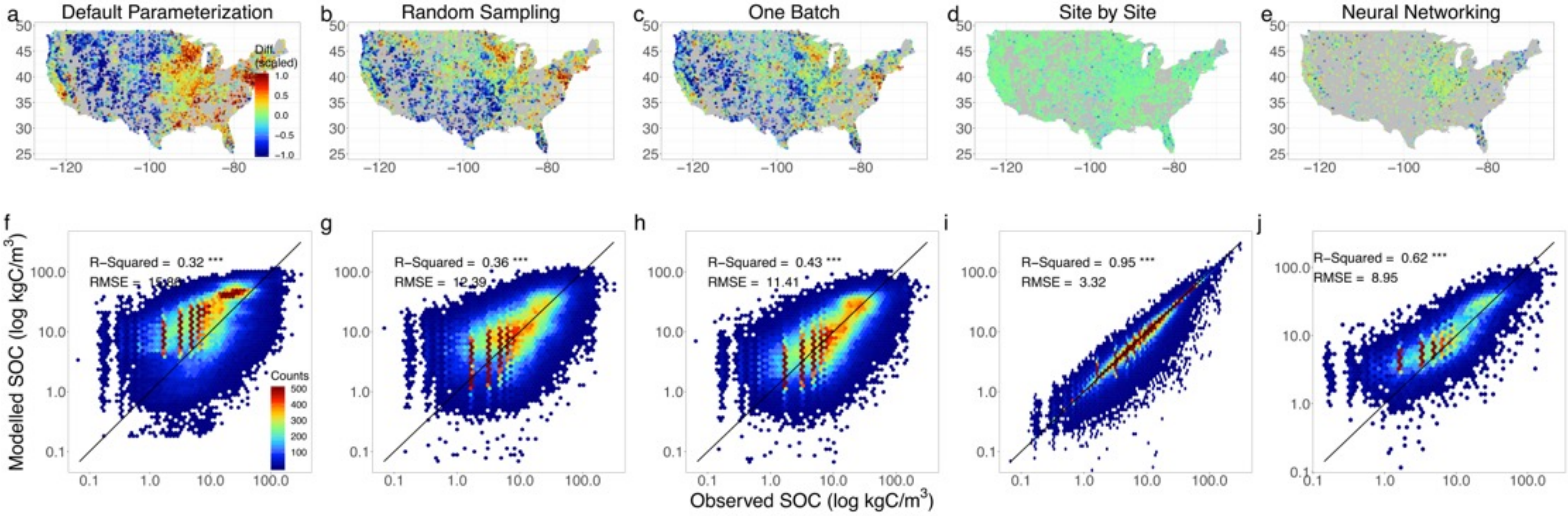
Trained parameters

Machine learning





Tao et al. *2020*
Frontier in Big Data



Tao et al. *2020*

Frontier in Big Data

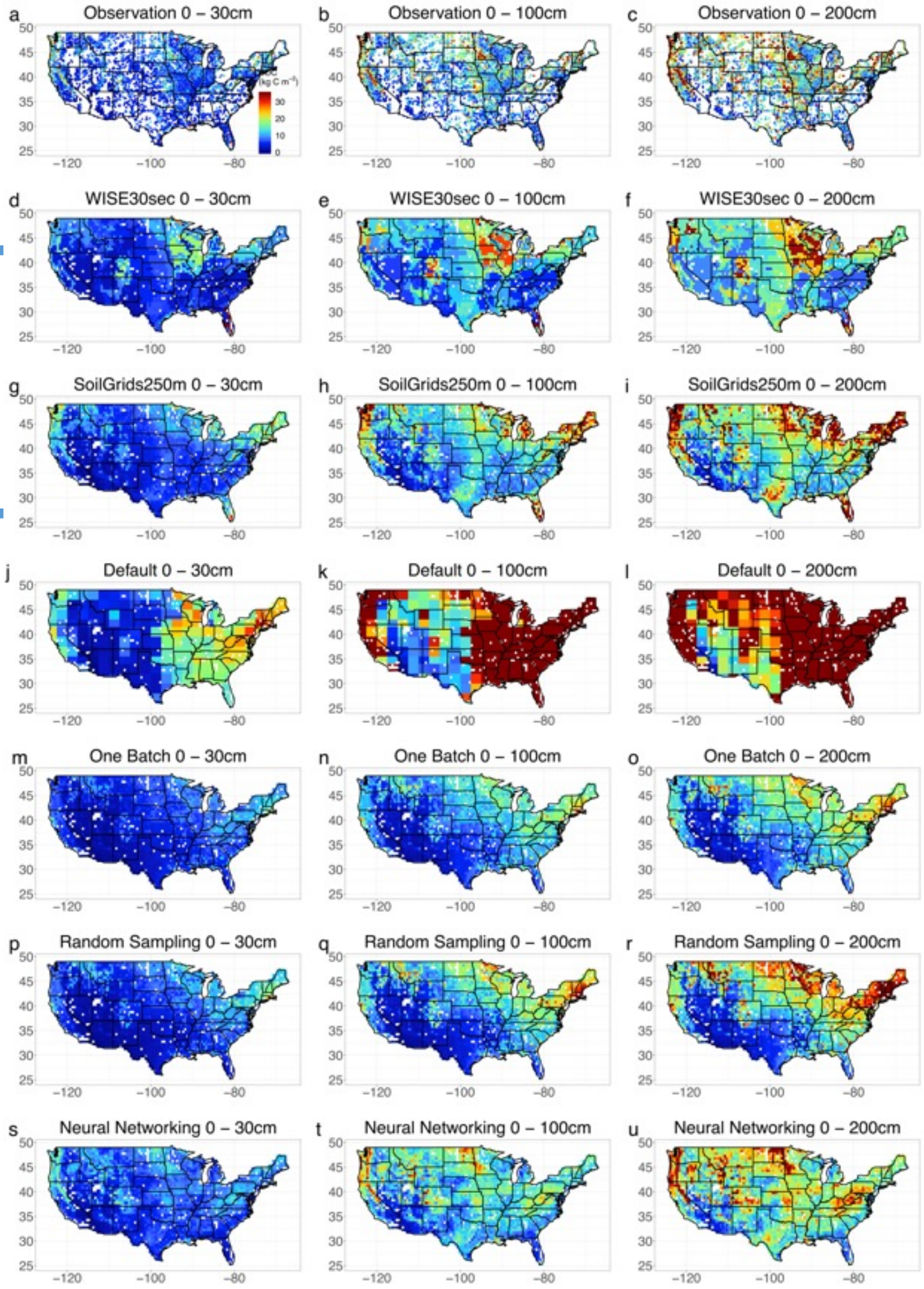
Observations

Data products

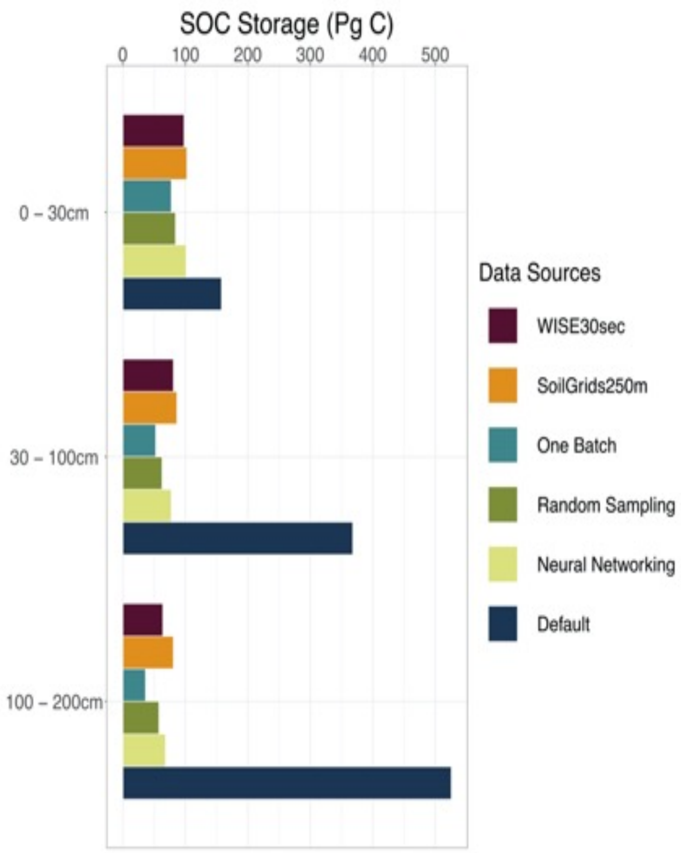
Model with default parameters

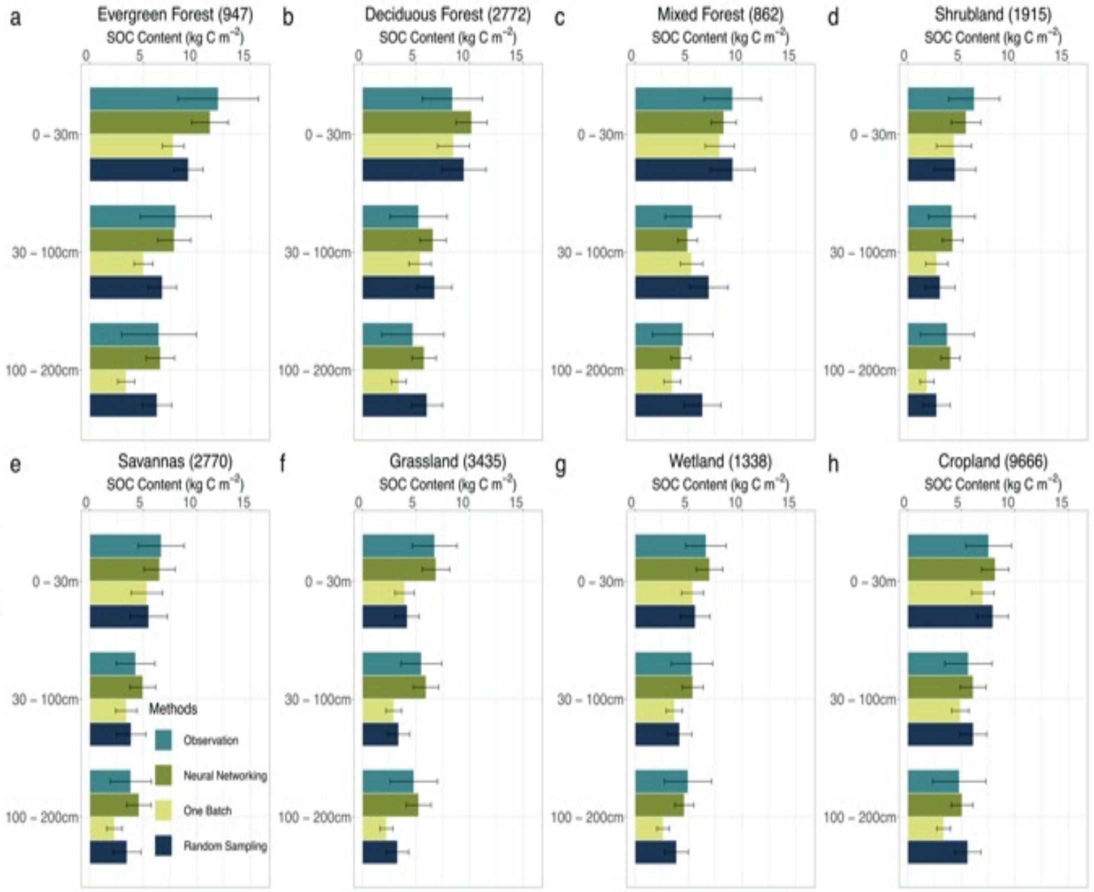
Data-driven model predictions

Combined deep learning and model predictions



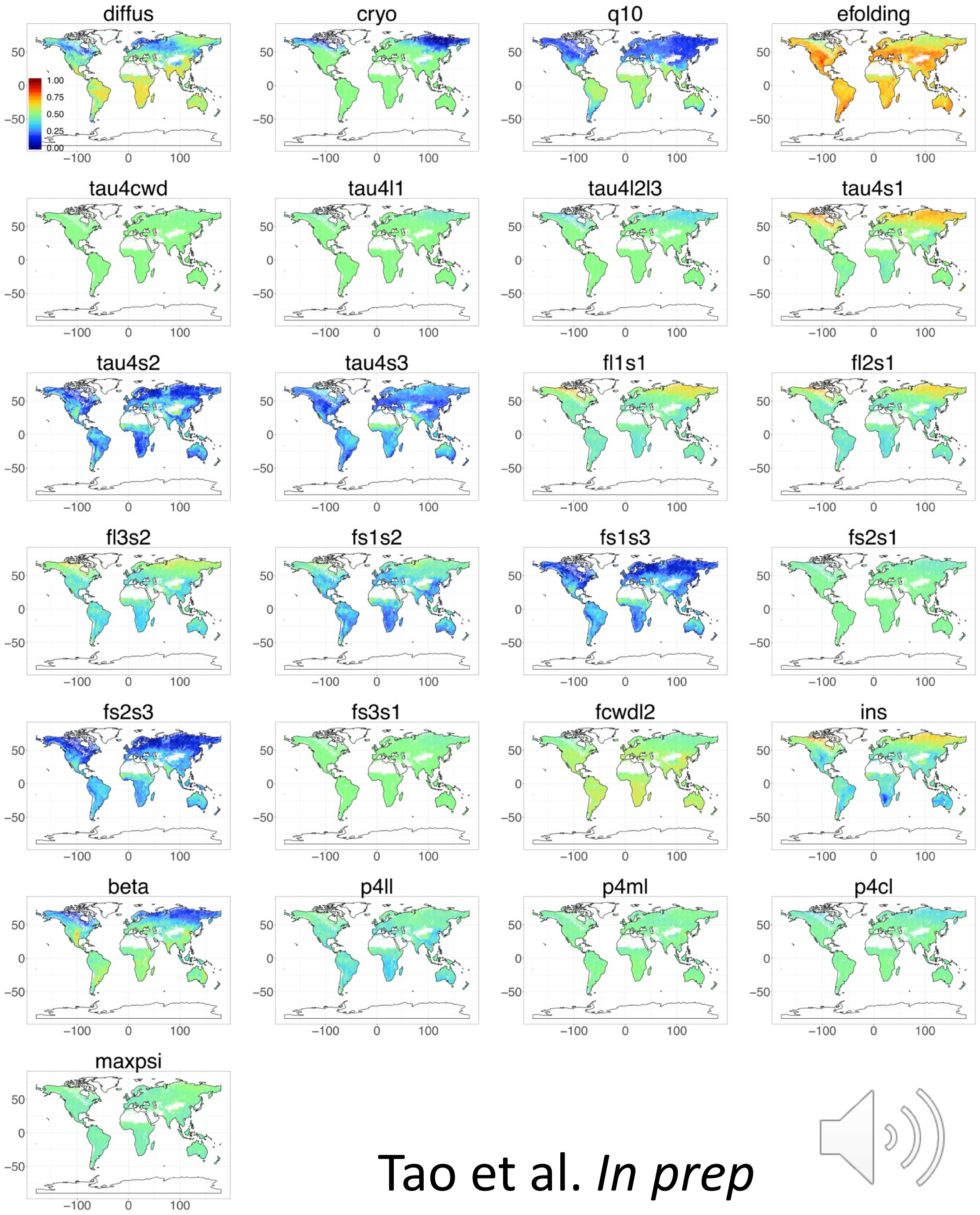
Tao et al. *2020*
Frontier in Big Data



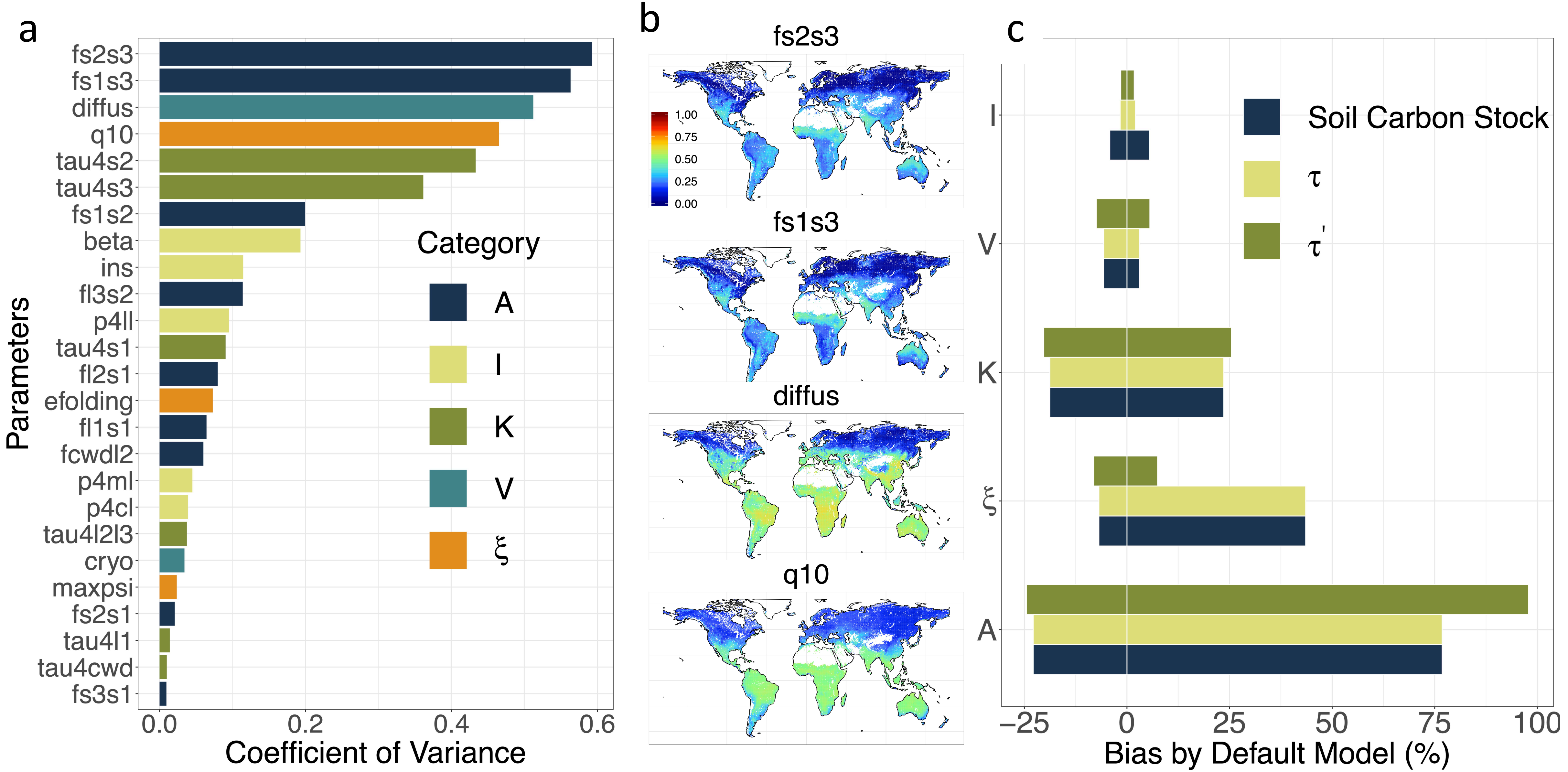


Tao et al. *2020*
Frontier in Big Data

Patterns of spatially heterogeneous parameters



Tao et al. *In prep*



Parameters in category *A* are related to microbial carbon use efficiency (CUE_m). Results indicate that CUE_m has to vary in order to predict soil carbon well.

Tao et al. *In prep*

Conclusions

- ✓ Spatially heterogeneous mechanisms are required to realistically predict states of soil carbon dynamics
- ✓ Big data offer new opportunities to discover such mechanisms
- ✓ We developed a novel approach – combined PROcess-guided machine learning and Data-driven modeling (PRODA) – to uncover spatially heterogeneous mechanisms underlying soil carbon sequestration

The 1st and 2nd Training Courses on

New Advances in Land Carbon Cycle Modeling

Flagstaff, AZ, USA, 2018 and 2019

The 3rd Training Courses on

New Advances in Land Carbon Cycle Modeling

Virtual, 20-31 July 2020

http://www2.nau.edu/luo-lab/?workshop

1. New theory on land carbon storage dynamics
2. Matrix approach to land carbon, nitrogen, and phosphorus modeling
3. Data assimilation system with both flux- and pool-based observations
4. Deep learning and machine learning to enhance process-based research
5. Ecological forecasting